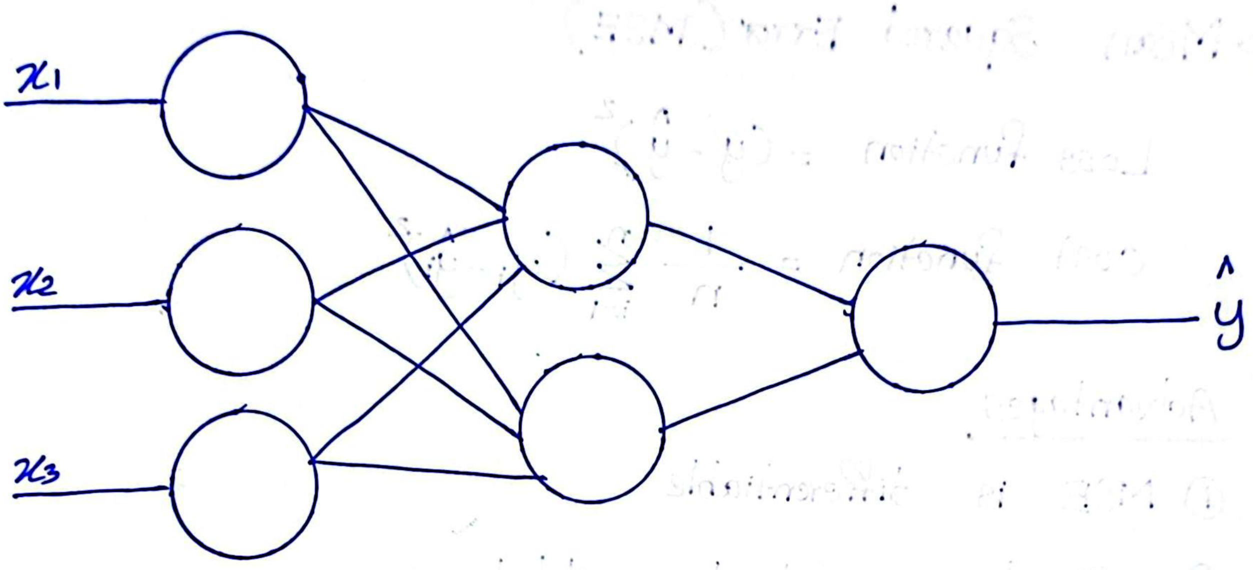
Loss and Cost Function



x_1	x_2	x_3	output
–	–	–	0
–	–	–	1
–	–	–	0
–	–	–	1

$$\text{loss} = (y-\hat{y})^2$$

$$\text{cost} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

Error/Metric	Loss function	Cost function
MSE	$(y-\hat{y})^2$	$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

→ Regression Cost Function

> Mean Squared Error (MSE)

Loss function $= (y - \hat{y})^2$

cost function $= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

Advantages

1. MSE is differentiable
2. MSE has one local or global minima
3. MSE converges faster

Disadvantages

1. Not robust to outliers

> Mean Absolute Error (MAE)

Loss function $= |y - \hat{y}|$

Cost function $= \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$

Advantages

1. MAE is robust to outliers

Disadvantages

1. MAE converges slower

> Huber Loss

Combination of MAE and MSE

$$\text{Cost function} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 & \text{if } |y-\hat{y}| \le \delta \rightarrow \text{hyperparameter} \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases}$$

> Root Mean Squared Error (RMSE)

$$\text{Loss function} = \sqrt{(y-\hat{y})^2}$$

$$\text{Cost function} = \sqrt{\frac{\sum_{i=1}^{N}(y_i-y_i)^2}{N}}$$

Advantages

① Measure of error is in the same unit as the target variable

② Easier to interpret and compare

Disadvantages

① RMSE is sensitive to outliers

→ Classification Cost Function

> Binary Cross Entropy

It is used for binary classification

$$\text{loss} = -y\log\hat{y} - (1-y)\log(1-\hat{y})$$

$$\text{loss} = \begin{cases} -\log(1-\hat{y}) & \text{if } y=0 \\ -\log\hat{y} & \text{if } y=1 \end{cases}$$

> Categorical Cross Entropy

It is used for multiclass classification

f_1	f_2	f_3	output
–	–	–	Good
–	–	–	Bad
–	–	–	Neutral

⇓ one hot encoding

f_1	f_2	f_3	Good	Bad	Neutral
–	–	–	1	0	0
–	–	–	0	1	0
–	–	–	0	0	1

$$\mathcal{L}(x_i, y_i) = -\sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij})$$

C = no of categories

$$y_{ij} = \begin{cases} 1 & \text{if element is in class} \\ 0, & \text{otherwise} \end{cases}$$

> Sparse Categorical Cross Entropy

Suppose,

output of categorical cross entropy

[0.2, 0.3, 0.5]

By using sparse categorical cross entropy

[0.2, 0.3, 0.5] => 2nd Index

Final output

→ Which Loss Function to Use When

	Hidden layer	output layer	problem statement	loss function
①	ReLU	Sigmoid	Binary classification	Binary Cross Entropy
②	ReLU	Softmax	Multiclass Classification	Categorical or Sparse Categorical Entropy
③	ReLU	Linear	Regression	MSE, MAE, Huber loss or RMSE